COMPOSITIONS WHICH CAN BE USED FOR REGULATING THE ACTIVITY OF PARKIN

The present invention relates to compositions and methods which can be used for regulating the activity of parkin. It relates in particular to a novel protein, referred to as PAP1, which is a partner of parkin, as well as to the peptides or polypeptides which are derived from or are homologous to this protein. It also relates to compounds which are capable of modulating, at least partially, the activity of parkin, in particular of interfering with the interaction between parkin and PAP1. The present invention can be used in the therapeutic or diagnostic areas, or for forming pharmacological targets which make possible the development of novel drugs.

The parkin gene is mutated in certain familial forms (autosomal recessive juvenile) of Parkinson's disease (Kitada *et al.*, 1998). Parkinson's disease (Lewy, 1912) is one of the most common neurodegenerative diseases, affecting more than 1% of the population over 55 years old. Patients suffering from this disease have neurological disorders which are grouped together under the term "Parkinsonian syndrome," which is characterized by rigidity, bradykinesia, and resting tremor. These symptoms are the consequence of a degeneration of the dopaminergic neurons of the substantia nigra of the brain.

Most cases with a Parkinson's disease do not have a familial history. However, familial cases do exist, of which certain correspond to a monogenic form of the disease. At the current time, only three different genes have been identified in certain rare hereditary forms. The first form corresponds to an autosomal dominant form, in which the gene responsible encodes alpha Synuclein (Polymeropoulos *et al.*, 1997). This protein is an abundant constituent of the intracytoplasmic inclusions, termed Lewy bodies, which are used as a marker for Parkinson's disease (Lewy, 1912). The second form, also autosomal dominant, is linked to a mutation in a gene which encodes a hydrolase termed ubiquitin carboxy-terminal hydrolase L1 (Leroy *et al.*, 1998). This

enzyme is assumed to hydrolyze ubiquitin polymers or conjugates into ubiquitin monomers. The third form differs from the previous forms in that it has an autosomal recessive transmission and onset which often occurs before 40 years of age, as well as an absence of Lewy bodies. These patients respond more favorably to levodopa, a dopamine precursor which is used as treatment for Parkinson's disease. The gene involved in this form encodes a novel protein which is termed parkin (Kitada *et al.*, 1998).

The parkin gene consists of 12 exons which cover a genomic region of more than 500,000 base pairs on chromosome 6 (6q25.2-q27). At the current time, two major types of mutation of this gene, which are at the origin of the disease, are known; either deletions of variable size in the region which covers exons 2 to 9, or point mutations which produce the premature appearance of a stop codon or the change of an amino acid (Kitada *et al.*, 1998; Abbas *et al.*, 1999; Lucking *et al.*, 1998; Hattori *et al.*, 1998). The nature of these mutations and the autosomal recessive method of transmission suggest a loss of function of the parkin, which leads to Parkinson's disease.

This gene is expressed in a large number of tissues and in particular in the substantia nigra. Several transcripts which correspond to this gene and originate from different alternative splicing sites Kitada *et al.*, 1998; Sunada *et al.*, 1998) exist. In the brain, two types of messenger RNAs are found, of which one lacks the portion corresponding to exon 5. In the leukocytes, parkin messenger RNAs which do not contain the region encoding exons 3, 4 and 5 have been identified. The longest of the parkin messenger RNAs, which is present in the brain, contains 2960 bases and encodes a protein of 465 amino acids.

This protein has a slight homology with ubiquitin in its N-terminal portion. Its C-terminal half contains two ring finger motifs, separated by an IBR (In Between Ring) domain, which correspond to a cysteine-rich region and which are able

to bind metals, like the zinc finger domains (Morett, 1999). It has been shown by immunocytochemistry that parkin is located in the cytoplasm and the Golgi apparatus of neurons of the substantia nigra which contain melanin (Shimura *et al.*, 1999). In addition, this protein is present in certain Lewy bodies of Parkinsonians. The cellular function of parkin has not yet been demonstrated, but it might play a transporter role in synaptic vesicles, in the maturation or degradation of proteins, and in the control of cellular growth, differentiation or development. In the autosomal recessive juvenile forms, parkin is absent, which thus confirms that the loss of this function is responsible for the disease.

The elucidation of the exact role of the parkin protein in the process of degeneration of the dopaminergic neurons thus constitutes a major asset for the understanding of and the therapeutic approach to Parkinson's disease, and more generally diseases of the central nervous system.

The present invention lies in the identification of a partner of parkin, which interacts with this protein under physiological conditions. This partner represents a novel pharmacological target for manufacturing or investigating compounds which are capable of modulating the activity of parkin, in particular its activity on the degeneration of dopaminergic neurons and/or the development of nervous pathologies. This protein, the antibodies, the corresponding nucleic acids, as well as the specific probes or primers, can also be used for detecting or assaying the proteins in biological samples, in particular nervous tissue samples. These proteins or nucleic acids can also be used in therapeutic approaches, for modulating the activity of parkin and any compound according to the invention which is capable of modulating the interaction between parkin and the polypeptides of the invention.

The present invention results more particularly from the demonstration by the applicant of a novel human protein, referred to as PAP1 (Parkin

Associated Protein 1), or LY111, which interacts with parkin. The PAP1 protein (sequence SEQ ID NO: 1 or 2) shows a certain homology with synaptotagmins and is capable of interacting more particularly with the central region of parkin (represented on the sequence SEQ ID NO: 3 or 4). The PAP1 protein has also been cloned, sequenced and characterized from various tissues of human origin, specifically lung (SEQ ID NO: 12, 13) and brain (SEQ ID NO: 42, 43) tissue, as well as short forms, which correspond to splicing variants (SEQ ID NO: 14, 15, 44, 45).

The present invention also results from the identification and characterization of specific regions of the PAP1 protein which are involved in the modulation of the function of parkin. The demonstration of the existence of this protein and of regions which are involved in its function makes it possible in particular to prepare novel compounds and/or compositions which can be used as pharmaceutical agents, and to develop industrial methods of screening such compounds.

A first subject of the invention thus relates to compounds which are capable of modulating, at least partially, the interaction between the PAP1 protein (or homologs thereof) and parkin (in particular human parkin), or of interfering with the interaction between these proteins.

Another subject of the invention lies in the PAP1 protein and fragments, derivatives and homologs thereof.

Another aspect of the invention lies in a nucleic acid which encodes the PAP1 protein or fragments, derivatives or homologs thereof, as well as any vector which comprises such a nucleic acid and any recombinant cell which contains such a nucleic acid or vector, and any non-human mammal comprising such a nucleic acid in its cells.

The invention also relates to antibodies which are capable of binding the PAP1 protein and fragments, derivatives and homologs thereof, in particular polyclonal or monoclonal antibodies, more preferably antibodies which are capable of binding the PAP1 protein and of inhibiting, at least partially, its interaction with parkin.

Another aspect of the invention relates to nucleotide probes or primers, which are specific to PAP1 and which can be used for detecting or amplifying the PAP1 gene, or a region of this gene, in any biological sample.

The invention also relates to pharmaceutical compositions, methods for detecting genetic abnormalities, methods for producing polypeptides as defined above and methods for screening or for characterizing active compounds.

As indicated above, a first aspect of the invention lies in a compound which is capable of interfering, at least partially, with the interaction between the PAP1 protein (or homologs thereof) and parkin.

For the purposes of the present invention, the name PAP1 protein refers to the protein per se, as well as to all homologous forms thereof. "Homologous form" is intended to refer to any protein which is equivalent to the protein under consideration, of varied cellular origin and in particular derived from cells of human origin, or from other organisms, and which possesses an activity of the same type. Such homologs also comprise natural variants of the PAP1 protein of sequence SEQ ID NO 2, in particular polymorphic or splicing variants. Such homologs can be obtained by experiments of hybridization between the coding nucleic acids (in particular the nucleic acid of sequence SEQ ID NO: 1). For the purposes of the invention, a sequence of this type only has to have a significant percentage of identity to lead to a physiological behavior which is comparable to that of the PAP1 protein as claimed. "Significant percentage of identity" is intended to refer to a percentage of at least 60%, preferably 80%, more preferably 90% and even more preferably 95%. As such, variants and/or homologs of the sequence SEQ ID NO: 2 are described in the sequences SEQ ID NO: 13, 15, 43 and 45, and are identified from tissues of human origin. The name PAP1 therefore also encompasses these polypeptides.

For the purposes of the present invention, the "percentage of identity" between two sequences of nucleotides or amino acids can be determined by comparing two optimally aligned sequences through a window of comparison.

The part of the nucleotide or polypeptide sequence in the window of comparison can thus comprise additions or deletions (gaps, for example) as compared to the reference sequence (which does not contain these additions or deletions) such that an optimal alignment of the two sequences is obtained.

The percentage is calculated by determining the number of positions at which a nucleic acid base or identical amino acid residue is observed for the two sequences (nucleic acid or peptide) being compared, then dividing the number of positions at which there is identity between the two amino acid residues or bases by the total number of positions in the window of comparison, then multiplying the result by 100 so as to obtain the sequence identity percentage.

Optimal alignment of the sequences for purposes of the comparison can be performed on a computer using known algorithms contained in the Wisconsin Genetics Software Package, produced by Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin.

For purposes of illustration, the sequence identity percentage may be obtained with the BLAST software (BLAST versions 1.4.9 of March 1996, 2.0.4 of February 1998 and 2.0.6 of September 1998) using only the default parameters (Altschul et al, *J. Mol. Biol.* (1990) **215**:403-410; Altschul et al, *Nucleic Acids Res.* (1997) **25**: 3389-3402). Blast searches for sequences which are similar/homologous to a reference "query" sequence, using the Altschul et al algorithm (above). The query sequence and the databases used can be peptide or nucleic acid, with any combination being possible.

The interference of a compound according to the invention can reveal itself in various ways. Thus, the compound can slow, inhibit or stimulate, at least partially, the interaction between the PAP1 protein, or a homologous form thereof, and parkin. Preferably, they are compounds which are capable of modulating this

interaction in vitro, for example in a double-hybrid type system or in any acellular system for detecting an interaction between two polypeptides. The compounds according to the invention are preferably compounds which are capable of modulating, at least partially, this interaction, preferably of increasing or inhibiting this reaction by at least 20%, more preferably by at least 50%, as compared to a control in the absence of the compound.

In a particular embodiment, they are compounds which are capable of interfering with the interaction between the region of parkin which is represented on the sequence SEQ ID NO: 4 and the region of the PAP1 protein which is represented on the sequence SEQ ID NO: 2, 13, 15, 43 or 45.

According to a particular embodiment of the invention, the compounds are capable of binding at the domain of interaction between the PAP1 protein, or a homologous form thereof, and parkin.

The compounds according to the present invention can be varied in nature and in origin. In particular, they can be compounds of peptide, nucleic acid (i.e. comprising a string of bases, in particular a DNA or an RNA molecule), lipid or saccharide type, an antibody, etc. and, more generally, any organic or inorganic molecule.

According to a first variation, the compounds of the invention are peptide in nature. The term "peptide" refers to any molecule comprising a string of amino acids, such as for example a peptide, a polypeptide, a protein or an antibody (or antibody fragment or derivative), which, if necessary, is modified or combined with other compounds or chemical groups. In this respect, the term "peptide" refers more specifically to a molecule comprising a string of at most 50 amino acids, more preferably of at most 40 amino acids. A polypeptide (or a protein) preferably comprises from 50 to 500 amino acids, or more.

According to a first preferred embodiment, the compounds of the invention are peptide compounds comprising all or part of the peptide sequence SEQ ID NO: 2 or a derivative thereof, in particular all or part of the peptide sequence SEQ ID NO: 13, 15, 43 or 45 or derivatives of these sequences, more

particularly of the PAP1 protein, which comprises the sequence SEQ ID NO: 2, 13, 15, 43 or 45.

For the purposes of the present invention, the term "derivative" refers to any sequence which differs from the sequence under consideration because of a degeneracy of the genetic code, which is obtained by one or more modifications of genetic and/or chemical nature, as well as any peptide which is encoded by a sequence which hybridizes with the nucleic acid sequence SEQ ID NO: 1, or a fragment of this sequence, for example with the nucleic acid sequence SEQ ID NO: 12, 14, 42 or 44 or a fragment of these sequences, and which is capable of interfering with the interaction between the PAP1 protein, or a homolog thereof, and parkin. "Modification of genetic and/or chemical nature" can mean any mutation, substitution, deletion, addition and/or modification of one or more residues. The term "derivative" also comprises the sequences which are homologous to the sequence under consideration, which are derived from other cellular sources and in particular cells of human origin, or from other organisms, and which possess an activity of the same type. Such homologous sequences can be obtained by hybridization experiments. The hybridizations can be carried out with nucleic acid libraries, using the native sequence or a fragment of this sequence as probe, under varied conditions of hybridization (Maniatis *et al.*, 1989). Moreover, the term "fragment" or "part" refers to any portion of the molecule under consideration, which comprises at least 5 consecutive residues, preferably at least 9 consecutive residues, even more preferably at least 15 consecutive residues. Typical fragments can comprise at least 25 consecutive residues.

Such derivatives or fragments can be generated with different aims, such as in particular that of increasing their therapeutic effectiveness or of reducing their side effects, or that of conferring novel pharmacokinetic and/or biological properties thereon.

As a peptide which is derived from the PAP1 protein and from the homologous forms, mention may be made in particular of any peptide which is capable of interacting with parkin, but which bears an effector region which has been made nonfunctional. Such peptides can be obtained

by deletion, mutation or disruption of this effector region on the PAP1 protein and homologous forms. Such modifications can be carried out for example by *in vitro* mutagenesis, by introducing additional elements or synthetic sequences, or by deletions or substitutions of the original elements. When such a derivative as defined above is prepared, its activity as partial inhibitor of the binding of the PAP1 protein, and of the homologous forms on its binding site on parkin, can be demonstrated. Any technique known to one skilled in the art can of course be used for this purpose.

They can also be fragments of the sequences indicated above. Such fragments can be generated in various ways. In particular they can be synthesized chemically, on the basis of the sequences given in the present application, using the peptide synthesizers known to one skilled in the art. They can also be synthesized genetically, by expression in a host cell of a nucleotide sequence which encodes the desired peptide. In this case, the nucleotide sequence can be prepared chemically using an oligonucleotide synthesizer, on the basis of the peptide sequence given in the present application and of the genetic code. The nucleotide sequence can also be prepared from sequences given in the present application, by enzymatic cleavage, ligation, cloning, etc., according to the techniques known to one skilled in the art, or by screening DNA libraries with probes which are developed from these sequences.

Moreover, the peptides of the invention, i.e., which are capable of modulating, at least partially, the interaction between the PAP1 protein, and homologous forms, and parkin, can also be peptides which have a sequence corresponding to the site of interaction of the PAP1 protein and of the homologous forms on parkin.

Other peptides according to the invention are peptides which are capable of competing with the peptides defined above for the interaction with their cellular

target. Such peptides can be synthesized in particular on the basis of the sequence of the peptide under consideration, and their capacity for competing with the peptides defined above can be determined.

A specific subject of the present invention relates to the PAP1 protein. It is more particularly the PAP1 protein comprising the sequence SEQ ID NO: 2 or a fragment or derivative of this sequence, for example the PAP1 protein, sequence SEQ ID NO: 13, 15, 43, 45 or fragments of these sequences.

Another subject of the invention lies in polyclonal or monoclonal antibodies or antibody fragments or derivatives, which are directed against a polypeptide as defined above. Such antibodies can be generated by methods known to one skilled in the art. In particular, these antibodies can be prepared by immunizing an animal against a peptide compound of the invention (in particular a polypeptide or a peptide comprising all or part of the sequence SEQ ID NO: 2), sampling the blood and isolating the antibodies. These antibodies can also be generated by preparing hybridomas according to the techniques known to one skilled in the art.

More preferably, the antibodies or antibody fragments of the invention have the capacity to modulate, at least partially, the interaction of the claimed peptides with parkin.

Moreover, these antibodies can also be used for detecting and/or assaying the expression of PAP1 in biological samples and, consequently, for providing information on its activation state.

The antibody fragments or derivatives are for example Fab or Fab'2 fragments, single-chain antibodies (ScFv), etc. They are in particular any fragment or derivative which retains the antigenic specificity of the antibodies from which they are derived.

The antibodies according to the invention are more preferably capable of binding the PAP1 proteins which comprise the sequence SEQ ID NO: 2, 13, 43 or 45, in particular

the region of this protein which is involved in the interaction with parkin. These antibodies (or fragments or derivatives) are more preferably capable of binding an epitope which is present in the sequence between residues 1 and 344 of the sequence SEQ ID NO: 2.

The invention also relates to compounds which are not peptide or not exclusively peptide, which can be used as a pharmaceutical agent. It is in fact possible, from the active protein motifs described in the present application, to prepare molecules which are modulators of the activity of PAP1, are not exclusively peptide, and are compatible with pharmaceutical use, in particular by duplicating the active motifs of the peptides with a structure which is not a peptide, or which is not of exclusively peptide nature.

A subject of the present invention is also any nucleic acid which encodes a peptide compound according to the invention. It can be, in particular, a nucleic acid comprising all or part of the sequence which is presented in SEQ ID NO: 1, 12, 14, 42 or 44 or a derivative thereof. For the purposes of the present invention, "derived sequence" is intended to mean any sequence which hybridizes with the sequence which is presented in SEQ ID NO: 1, or with a fragment of this sequence, and which encodes a peptide compound according to the invention, as well as the sequences which result from the latter by degeneracy of the genetic code. For example, nucleic acids according to the invention comprise all or part of the nucleic sequence SEQ ID NO: 12, 14, 42 or 44.

Moreover, the present invention relates to sequences which have a significant percentage of identity with the sequence presented in SEQ ID NO: 1 or with a fragment thereof and which encodes a peptide compound with physiological behavior which is comparable to that of the PAP1 protein. "Significant percentage of identity" is intended to mean a percentage of at least 60%, preferably 80%, more preferably 90% and even more preferably 95%.

The various nucleotide sequences of the invention may or may not be of artificial origin. They can be genomic, cDNA or RNA sequences,

hybrid sequences or synthetic or semi-synthetic sequences. These sequences can be obtained either by screening DNA libraries (cDNA library, genomic DNA library), by chemical synthesis, by mixed methods which include the chemical or enzymatic modification of sequences which are obtained by screening of libraries, or by searching for homology in nucleic acid or protein databases. The abovementioned hybridization is preferably carried out under the conditions described by Sambrook et al. (1989, pages 9.52-9.55).

It is advantageously carried out under highly stringent hybridization conditions. For the purposes of the present invention, "highly stringent hybridization conditions" is intended to mean the following conditions:

1- Competition of the membranes and PRE-HYBRIDIZATION:

- Mix: 40µl salmon sperm DNA (10 mg/ml)
  + 40µl human placenta DNA (10 mg/ml)
- Denature for 5 min. at 96°C, then immerse the mixture in ice.
- Remove the SSC 2X buffer and pour 4 ml formamide mix into the hybridization tube which contains the membranes.
- Add the mixture of the two denatured DNAs.
- Incubate at 42°C for 5 to 6 hours, with rotation.

2-Competition of the labeled probe:

- Add to the labeled and purified probe 10 to 50µl Cot I DNA, according to the quantity of non-specific hybridizations.
- Denature 7 to 10 min. at 95°C.
- Incubate at 65°C for 2 to 5 hours.

3-Hybridization:

- Remove the pre-hybridization mix
- Mix 40 µl salmon sperm DNA + 40 µl human placenta DNA; denature 5 min. at 96°C, then immerse in ice.

- Add to the hybridization tube 4 ml formamide mix, the mixture of the two DNAs and the labeled probe/denatured Cot I DNA.
- Incubate 15 to 20 hours at 42°C, with rotation.

4-Washes:

- One wash at room temperature in SSC 2X, to rinse.
- 2 times 5 minutes at room temperature SSC 2X and SDS 0.1%.
- 2 times 15 minutes SSC 0.1X and SDS 0.1% at 65°C

Wrap membranes in Saran and expose.

The hybridization conditions described above are suitable for hybridization under highly stringent conditions of a nucleic acid molecule varying in length from 20 nucleotides to several hundred nucleotides.

The hybridization conditions described above could of course be adjusted to take into account the length of the nucleic acid for which hybridization is desired or the type of label chosen, according to techniques known to one skilled in the art.

For example, the suitable hybridization conditions can be adjusted according to the teachings contained in the work of Hames and Higgins (1985) (Nucleic Acid Hybridization a Practical Approach, Hames and Higgins Ed., IRL Press, Oxford) or, alternatively, in the work of F. Ausubel et al (1999) (Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, NY).

For the purposes of the invention, a particular nucleic acid encodes a polypeptide comprising the sequence SEQ ID NO: 2 or a fragment or derivative of this sequence, in particular the human PAP1 protein. It is advantageously a nucleic acid comprising the sequence SEQ ID NO: 1, 12, 14, 42 or 44.

Such nucleic acids can be used for producing the peptide compounds of the invention. The present application thus relates to a method for preparing such peptide compounds, according to which a cell which contains a nucleic acid according to the invention is cultured under conditions for expressing said

nucleic acid, and the peptide compound produced is recovered. In this case, the portion which encodes said peptide compound is generally placed under the control of signals which allow its expression in a host cell. The choice of these signals (promoters, terminators, secretion leader sequence, etc.) can vary as a function of the host cell used. Moreover, the nucleic acids of the invention can form part of a vector which can replicate autonomously or can integrate. More particularly, autonomously-replicating vectors can be prepared using sequences which replicate autonomously in the chosen host. As regards the integrating vectors, they can be prepared for example using sequences which are homologous to certain regions of the genome of the host, which allow, by homologous recombination, the integration of the vector. It can be a vector of plasmid, episomal, chromosomal, viral etc., type.

The host cells which can be used for producing the peptide compounds of the invention via the recombinant pathway are both eukaryotic and prokaryotic hosts. Among the eukaryotic hosts which are suitable, mention may be made of animal cells, yeasts or fungi. In particular, as regards yeasts, mention may be made of the yeasts of the genus *Saccharomyces*, *Kluyveromyces*, *Pichia*, *Schwanniomyces*, or *Hansenula*. As regards animal cells, mention may be made of COS, CHO, C127, PC12 etc., cells. Among the fungi, mention may be made more particularly of *Aspergillus* ssp. or *Trichoderma* ssp. As prokaryotic hosts, use of the following bacteria is preferred: *E. coli*, *Bacillus* or *Streptomyces*.

A subject of the present invention is also non-human mammals comprising in their cells a nucleic acid or vector according to the invention.

Such mammals (rodents, canines, rabbits, etc.) can be used in particular to study the properties of PAP1 and identify compounds with therapeutic aims. The genome of such a transgenic animal can be modified by knock-in or knock-out alteration or modification of one or more genes. This modification can be carried out using conventional alterative or mutagenic agents, or via directed mutagenesis. Modification of the genome can also be the result of the insertion of a gene(s) or the replacement of a gene(s)

in its (their) wild or mutated form. Genome modifications are advantageously carried out on reproductive stem cells and advantageously on pronuclei. Transgenesis can be performed by microinjection of an expression cassette comprising the modified genes in the two fertile pronuclei. Thus an animal according to the present invention can be obtained by injection of an expression cassette comprising a nucleic acid. Preferably, this nucleic acid is a DNA which can be a genomic DNA (gDNA) or a complementary DNA (cDNA).

The construction of transgenic animals according to the invention can be carried out according to conventional techniques well known to one skilled in the art. A person skilled in the art can in particular refer to the production of transgenic animals, and specifically to the production of transgenic mice, as described in the following patents US 4,873,191; US 5,464,764 and US 5,789,215; the contents of these documents are incorporated herein by reference.

In short, a polynucleotide construct which comprises a nucleic acid according to the invention is inserted into an ES-type stem cell line. Insertion of the polynucleotide construct is preferably performed by electroporation, as described by Thomas et al. (1987, Cell, Vol. 51; 503-512).

The cells which have been subjected to the electroporation step are then screened for the presence of the polynucleotide construct (for example by selection, using markers, or alternatively by PCR or by Southern-type DNA gel electrophoresis analysis) so as to select the positive cells which integrated the exogenous polynucleotide construct into their genome, if necessary after a homologous recombination event. Such a technique is described by Mansour et al, for example. (*Nature* (1988) **336**: 348-352).

The positively selected cells are then isolated, cloned and injected into 3.5 day-old mouse blastocysts, as described by Bradley (1987, *Production and Analysis of Chimaeric Mice.* In: E.J. Robertson (Ed., Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, IRL Press. Oxford, page 113)). Blastocysts are then introduced

into a female animal host and development of the embryo is pursued to full term.

Alternatively, positively selected ES-type cells are placed in contact with 2.5 day-old embryos at an 8-16 cell stage (*morulae*), as described by Wood et al. (1993. Proc. Natl. Acad. Sci. USA, vol. 90: 4582-4585) or by Nagy et al. (1993. Proc. Natl. Acad. Sci. USA, vol. 90: 8424-8428). The ES cells are internalized in order to extensively colonize the blastocyst, including the cells which produce the germ line.

The descendants are then tested to determine those which have integrated the polynucleotide construct (the transgene).

The nucleic acids according to the invention can also be used to prepare genetic antisense or antisense oligonucleotides which can be used as pharmaceutical agents. Antisense sequences are oligonucleotides of short length, which are complementary to the coding strand of a given gene, and consequently are capable of specifically hybridizing with the mRNA transcript, which inhibits its translation into a protein. A subject of the invention is thus antisense sequences which are capable of inhibiting, at least partially, the interaction of the PAP1 proteins on parkin. Such sequences can consist of all or part of the nucleic acid sequences defined above. They are generally sequences, or fragments of sequences, which are complementary to sequences encoding peptides which interact with parkin. Such oligonucleotides can be obtained by fragmentation, etc., or by chemical synthesis.

The claimed sequences can be used in the context of gene therapies, for transferring and expressing, *in vivo*, antisense sequences or peptides which are capable of modulating the interaction of the PAP1 protein with parkin. In this respect, the sequences can be incorporated in viral or nonviral vectors, which allows their administration *in vivo* (Kahn *et al.*, 1991). As viral vectors in accordance with the invention, mention may be made most particularly of

adenovirus, retrovirus, adeno-associated virus (AAV) or herpes virus type vectors. A subject of the present application is also recombination-defective viruses comprising a nucleic acid which encodes a polypeptide according to the invention, in particular a polypeptide or peptide comprising all or part of the sequence SEQ ID NO: 2 or of a derivative of this sequence, for example all or part of the sequence SEQ ID NO: 12, 14, 42 or 44 or derivatives of these sequences.

The invention also enables the preparation of nucleotide probes, which may or may not be synthetic, and which are capable of hybridizing with the nucleotide sequences defined above or their complementary strand. Such probes can be used *in vitro* as a diagnostic tool for detecting the expression or overexpression of PAP1, or alternatively for revealing genetic abnormalities (incorrect splicing, polymorphism, point mutations, etc.). These probes can also be used for detecting and isolating homologous nucleic acid sequences which encode peptides as defined above, from other cellular sources and preferably from cells of human origin. The probes of the invention generally comprise at least 10 bases, and they can for example comprise up to the whole of one of the abovementioned sequences or of their complementary strand. Preferably, these probes are labeled prior to their use. For this, various techniques known to one skilled in the art can be employed (radioactive, fluorescent, enzymatic, chemical labeling, etc.).

The invention also relates to primers or primer pairs which make it possible to amplify all or part of a nucleic acid encoding a PAP1, for example a sequence primer chosen from among SEQ ID NO: 16-41.

A subject of the invention is also any pharmaceutical composition which comprises, as an active agent, at least one compound as defined above, in particular a peptide compound.

A subject of the invention is in particular any pharmaceutical composition which comprises, as an active agent, at least one antibody and/or one antibody fragment as

defined above, as well as any pharmaceutical composition which comprises, as an active agent at least one nucleic acid or one vector as defined above.

A subject of the invention is also any pharmaceutical composition which comprises, as an active agent, a chemical molecule which is capable of increasing or of decreasing the interaction between the PAP1 protein and parkin.

Moreover, a subject of the invention is also pharmaceutical compositions in which the peptides, antibodies, chemical molecules and nucleotide sequences defined above are combined mutually or with other active agents.

The pharmaceutical compositions according to the invention can be used for modulating the activity of the parkin protein, and consequently for maintaining the survival of the dopaminergic neurons. More particularly, these pharmaceutical compositions are intended for modulating the interaction between the PAP1 protein and parkin. They are, more preferably, pharmaceutical compositions which are intended for treating diseases of the central nervous system, such as for example Parkinson's disease.

A subject of the invention is also the use of the molecules described above for modulating the activity of parkin or for the typing of diseases of the central nervous system. In particular, the invention relates to the use of these molecules for modulating, at least partially, the activity of parkin.

The invention also relates to a method for screening or characterizing molecules which act on the function of parkin, to include selecting molecules which are capable of binding the sequence SEQ ID NO: 2 or the sequence SEQ ID NO: 4, or a fragment (or derivative) of these sequences. The method comprises, advantageously, bringing the molecule(s) to be tested into contact, in vitro, with a polypeptide which comprises the sequence SEQ ID NO: 2 or the sequence SEQ ID NO: 4, or a fragment (or derivative) of these sequences, and selecting molecules which are capable of binding the sequence SEQ ID NO: 2 (in particular the region between residues 1 and 344) or the sequence SEQ ID NO: 4. The molecules tested can be varied in nature

(peptide, nucleic acid, lipid, sugar, etc., or mixtures of such molecules, for example combinatory libraries, etc.). As indicated above, the molecules thus identified can be used for modulating the activity of the parkin protein, and represent potential therapeutic agents for treating neurodegenerative pathologies.

Other advantages of the present invention will appear upon reading the following examples and figure, which should be considered as illustrative and nonlimiting.

LEGENDS TO THE FIGURE:

Figure 1: Representation of the vector pLex9-parkin (135-290)

Figure 2: Results of the first 5'-RACE experiment. 8 clones were obtained. The initial sequence is indicated on the lower part of the figure.

Figure 3: Results of the second 5'-RACE experiment. Only two of the 8 clones obtained in the first experiment were validated (clones A12 and D5). The initial sequence is indicated on the lower part of the figure. The complete sequence of DNAs and proteins is provided in Sequences 12-15.

Figure 4: Detailed view of the organization of clones C5 and D4 from the second 5'-RACE experiment. The resulting consensus sequence is indicated on the upper part of the figure.

Figure 5: Structure of transcripts isolated from human brain.

Figure 6: LY111 (full length) nucleic acid and protein sequence from human brain. Double underlined: cysteines retained from zinc finger domain. Bold: Domain $C_2$1. Italics: domain $C_2$2.

Figure 7: LY111 (short version) nucleic acid and protein sequence from human brain. Double underlined: cysteines retained from zinc finger domain. Bold: Domain $C_2$1. Italics: domain $C_2$2.

Figure 8: Location of short (8b) or full length (8a) LY111 protein after expression in Cos-7 cells.

Figure 9: LY111 (full length) nucleic acid and protein sequence from human lung.

Figure 10: LY111 (short version) nucleic acid and protein sequence from human brain.

MATERIALS AND TECHNIQUES USED

1) Yeast strains:

Strain L40 of the genus *S. cerevisiae (Mata, his3D200, trp1-901, leu2-3, 112, ade2, LYS2:: (lexAop)$_4$-HIS3, URA3::(lexAop)$_8$-LacZ, GAL4, GAL80)* was used to verify the protein-protein interactions when one of the protein partners is fused to the LexA protein. The LexA protein is capable of recognizing the LexA response element, which controls the expression of the reporter genes *LacZ* and His3.

It was cultured on the following culture media:

Complete YPD medium: - Yeast extract (10 g/l) (Difco)
- Bactopeptone (20 g/l) (Difco)
- Glucose (20 g/l) (Merck)

This medium was solidified by addition of 20 g/l of agar (Difco).

Minimum YNB medium: - Yeast Nitrogen Base (without amino acids) (6.7 g/l) (Difco)
- Glucose (20 g/l) (Merck)

This medium can be solidified by addition of 20 g/l of agar (Difco). It can also be supplemented with amino acids and/or with 3-amino-1,2,4-triazole by addition of CSM media [CSM-Leu, -Trp, -His (620 mg/l), CSM-Trp (740 mg/l) or CSM-Leu, -Trp (640 mg/l)(Bio101)] and/or of 2.5 mM 3-amino-1,2,4-triazole.

2) Bacterial strains:

Strain TG1 of *Escherichia coli*, of genotype supE, hsdΔ5, thi, Δ(lac-proAB), F'[tra D36 pro A^+B^+ $lacI^q$ lacZΔM15], was used for constructing

plasmids, as a means of amplifying and of isolating recombinant plasmids used. It was cultured on the following medium:

Medium LB: - NaCl (5g/l) (Prolabo)
- Bactotryptone (10 g/l) (Difco)
- Yeast extract (5 g/l) (Difco)

This medium is solidified by addition of 15 g/l of agar (Difco).

Ampicillin was used at 100 µg/ml; this antibiotic is used to select the bacteria, which have received the plasmids bearing the gene for resistance to this antibiotic, as a marker.

Strain HB101 of *Escherichia coli* of genotype supE44, ara14, galK2, lacY1, Δ(gpt-proA)62, rpsL20(Str^r), xyl-5, mtl-1, recA13, Δ(mcrC-mrr), $HsdS^-(r^-m^-)$ was used as means for amplifying and isolating plasmids which originate from the human lymphocyte cDNA library.

It was cultured on

Medium M9: -Na_2HPO_4 (7 g/l) (Prolabo)
-KH_2PO_4 (3 g/l) (Prolabo)
-NH_4Cl (1 g/l) (Prolabo)
-NaCl (0.5 g/l) (Prolabo)
-Glucose (20 g/l) (Sigma)
-$MgSO_4$ (1 mM) (Prolabo)
-Thiamine (0.001%) (Sigma)

This medium is solidified by addition of 15 g/l of agar (Difco).

Leucine (50 mg/l) (Sigma) and proline (50 mg/l) (Sigma) should be added to the M9 medium to enable the growth of strain HB101.

During the selection of plasmids which originate from the lymphocyte cDNA two-hybrid library, leucine was not added to the medium because the plasmids bear a Leu2 selection marker.

3) Plasmids:

The 5-kb vector pLex9 (pBTM116) (Bartel *et al.*, 1993), which is homologous to pGBT10 and which contains a multiple cloning site located downstream of the sequence which encodes the LexA bacterial repressor, and upstream of a terminator, for forming a fusion protein.

pLex-HaRasVal12; plasmid pLex9, as described in application WO 98/21327, which contains the sequence encoding the HaRas protein mutated at position Val12, which is known to interact with the mammalian Raf protein (Vojtek *et al.*, 1993). This plasmid was used to test the specificity of interaction of the PAP1 protein in strain L40.

pLex9-cAPP; plasmid pLex9 which contains the sequence encoding the cytoplasmic domain of the APP protein, known to interact with the PTB2 domain of FE65. This plasmid was used to test the specificity of interaction of the PAP1 protein in strain L40.

4) Synthetic oligonucleotides:

TTAAGAATTC GGAAGTCCAG CAGGTAG	(SEQ ID N°5)
ATTAGGATCC CTACACACAA GGCAGGGAG	(SEQ ID N°6)

Oligonucleotides which made it possible to obtain the PCR fragment which corresponds to the central region of parkin, bordered by the *EcoRI* and *BamHI* sites.

GCGTTTGGAA TCACTACAG	(SEQ ID N°7)
GGTCTCGGTG TGGCATC	(SEQ ID N°8)
CCGCTTGCTT GGAGGAAC	(SEQ ID N°9)
CGTATTTCTC CGCCTTGG	(SEQ ID N°10)
AATAGCTCGA GTCAGTGCAG GACAAGAG	(SEQ ID N°11)

Oligonucleotides which were used to sequence the insert corresponding to the PAP1 gene.

The oligonucleotides are synthesized using an Applied System ABI 394-08 machine. They are removed from the synthesis matrix with ammonia and precipitated twice with 10 volumes of n-butanol, and then taken up in water. The

quantification is carried out by measuring the optical density (1 OD_{260} corresponds to 30 µg/ml).

5) Preparation of plasmid DNAs

The preparations of plasmid DNA were carried out according to the protocols recommended by Quiagen, the manufacturer of the DNA purification kits, in small and large amounts:

- Quiaprep Spin Miniprep kit, reference: 27106
- Quiaprep Plasmid Maxiprep kit, reference: 12613.

6) Enzymatic amplification of DNA by PCR (Polymerase Chain Reaction):

The PCR reactions are carried out in a final volume of 100 µl in the presence of the DNA matrix, of dNTP (0.2 mM), of PCR buffer (10 mM Tris-HCl pH 8.5, 1 mM $MgCl_2$, 5 mM KCl, 0.01% gelatin), of 10 to 20 pmol of each one of the oligonucleotides and of 2.5 IU of Ampli Taq DNA polymerase (Perkin Elmer). The mixture is covered with 2 drops of liquid petroleum jelly to limit the evaporation of the sample. The machine used is the "Crocodile II" by Appligene.

We used a matrix denaturation temperature of 94°C, a hybridization temperature of 52°C and a temperature for elongation by the enzyme at 72°C.

7) Ligations:

All the ligation reactions are carried out at 37°C for one hour in a final volume of 20 µl, in the presence of 100 to 200 ng of vector, 0.1 to 0.5 µg of insert, 40 IU of T4 DNA ligase enzyme (Biolabs) and a ligation buffer (50 mM Tris-HCl pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP). The negative control consists of ligating the vector in the absence of insert.

8) Transformation of bacteria:

The transformation of bacteria with a plasmid is carried out according to the following protocol: 10 µl of the ligation volume are used to transform the TG1 bacteria, according to the method of Chung (Chung *et al.*, 1989). After transformation, the bacteria are placed on an LB medium + ampicillin and incubated for 16 h at 37°C.

9) Separation and extraction of DNAs:

The separation of DNAs is carried out as a function of their size, on agarose gel by electrophoresis according to Maniatis (Maniatis *et al.*, 1989): 1% agarose gel (Gibco BRL) in a TBE buffer (90 mM Tris base; 90 mM borate; 2 mM EDTA).

10) Fluorescent sequencing of plasmid DNAs:

The sequencing technique used is derived from the method of Sanger (Sanger *et al.*, 1977) and adapted for sequencing by fluorescence, which is developed by Applied Biosystems. The protocol used is that described by the designers of the system (Perkin Elmer, 1997).

11) Transformation of yeast:

The plasmids are introduced into the yeast using a conventional technique for transforming yeast developed by Gietz (Gietz *et al.*, 1992) and modified in the following way:

In the specific case of the transformation of yeast with the lymphocyte cDNA library, the yeast used contains the plasmid pLex9-parkin (135-290), which encodes the central portion of parkin fused to the LexA protein. It is cultured in 200 ml of YNB minimum medium, supplemented with amino acids CSM-Trp, at 30°C with shaking until a density of 10^7 cells/ml is attained. To carry out the transformation of the yeasts, according to the above protocol, the cell suspension was separated into 10 50-µl tubes, into which 5 µg of the library were added. Heat shock was carried out for 20 minutes, and the cells were collected by centrifugation and resuspended in 100 ml of YPD medium for 1 h at 30°C, and

in 100 ml of YNB medium, supplemented with CSM-Leu, -Trp, for 3 h 30 at 30°C. The efficiency of the transformation is determined by placing various dilutions of transformed cells on solid YNB medium which is supplemented with CSM-Trp, -Leu. After 3 days of culture at 30°C, the colonies obtained were counted, and the rate of transformation per µg of lymphocyte library DNA was determined.

12) Isolation of plasmids extracted from yeast:

5 ml of a yeast culture, which is incubated for 16 h at 30°C, are centrifuged, and taken up in 200 µl of a lysis buffer (1M Sorbitol, 0.1 M KH_2PO_4/K_2HPO_4 pH 7.4, 12.5 mg/ml zymolyase) and incubated for 1 h at 37°C. The lysate is then treated according to the protocol recommended by Quiagen, the manufacturer of the DNA purification kit, Quiaprep Spin Miniprep kit, ref 27106.

13) β-galactosidase activity assay:

A sheet of nitrocellulose is preplaced on the Petri dish containing the yeast clones, which are separated from each other. This sheet is then immersed in liquid nitrogen for 30 seconds, in order to rupture the yeasts and thus to release the β-galactosidase activity. After thawing, the sheet of nitrocellulose is placed, colonies facing upwards, in another Petri dish containing a Whatman paper which has been presoaked in 1.5 ml of PBS solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7) containing 15 µl of X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactoside) at 40 mg/ml of N,N-dimethylformamide. The dish is then placed in an incubator at 37°C. The assay is termed positive when the colonies on the membrane turn blue after 12 hours.

EXAMPLE 1: CONSTRUCTION OF A VECTOR WHICH ALLOWS THE EXPRESSION OF A FUSION PROTEIN IN WHICH FUSION IS BETWEEN THE CENTRAL PORTION OF PARKIN AND THE LEXA BACTERIAL REPRESSOR.

Screening a library using the double-hybrid system requires the central region of parkin to be fused to a DNA binding protein, such as the LexA bacterial repressor. The expression of this fusion protein is carried out using the vector pLex9 (cf. materials and methods), into which the sequence encoding the central region of parkin, which is in the sequence presented in sequence SEQ ID NO: 3 or 4, is introduced, in the same reading frame as the sequence corresponding to the LexA protein.

The 468 bp-fragment of DNA corresponding to the 156 amino acids of the central region of parkin, which begins at amino acid 135, was obtained by PCR using the oligonucleotides (sequence SEQ ID NO: 5 and No. 6), which also made it possible to introduce the *EcoRI* site at the 5' end and a stop codon and a *BamHI* site at the 3' end. The PCR fragment was introduced between the *EcoRI* and *BamHI* sites of the multiple cloning site of the plasmid pLex9, downstream of the sequence encoding the protein LexA, in order to produce the vector pLex9-parkin (135-290) (Fig. 1).

The construct was verified by sequencing the DNA. This verification made it possible to show that this fragment does not have mutations generated during the PCR reaction, and that it was fused in the same open reading frame as that of the fragment corresponding to LexA.

EXAMPLE 2: SCREENING A LYMPHOCYTE FUSION LIBRARY

We used the double-hybrid method (Fields and Song, 1989).

Screening a fusion library makes it possible to identify clones producing proteins which are fused to the transactivating domain of GAL4, and which are able to interact with the protein of interest described in Example 1 (central region of parkin). This interaction makes it possible to reconstitute a transactivator which will then be capable of inducing the expression of the reporter genes His3 and *LacZ* in strain L40.

To carry out this screening we chose a fusion library which is prepared from cDNA originating from peripheral human lymphocytes, supplied by Richard

Benarous (Peytavi *et al.*, 1999). Yeasts were transformed with the lymphocyte library and positive clones were selected as described below.

During screening, it is necessary to maintain the probability that each separate plasmid from the fusion library is present in at least one yeast at the same time as the plasmid pLex9-parkin (135-290). To maintain this probability, it is important to have a good efficiency of transformation of the yeast. For this, we chose a protocol for transforming yeast which gives an efficiency of 2.6×10^5 transformed cells per µg of DNA. In addition, as cotransforming yeast with two different plasmids reduces this efficiency, we preferred to use a yeast which is pretransformed with the plasmid pLex9-parkin (135-290). This strain L40 pLex9-parkin (135-290), of phenotype His-, Lys-, Leu-, Ade-, was transformed with 50 µg of plasmid DNA from the fusion library. This amount of DNA enabled us to obtain, after estimation, 1.3×10^7 transformed cells, which corresponds to a number which is slightly higher than the number of separate plasmids which constitute the library. According to this result, virtually all of the plasmids of the library can be considered to have been used to transform the yeasts. The selection of the transformed cells, which are capable of reconstituting a functional transactivator, was done on a YNB medium which was supplemented with 2.5 mM 3-amino-1,2,4-triazole and 620 mg/l of CSM (Bio101), and which contains no histidine, no leucine and no tryptophan.

At the end of this selection, many clones with a His+ phenotype were obtained. A β-galactosidase activity assay was carried out on these transformants to validate, on the basis of the expression of the other reporter gene, *LacZ*, this number of obtained clones. 115 clones had the His+, β-Gal+ double phenotype, which can correspond to a protein-protein interaction.

EXAMPLE 3: ISOLATION OF THE LIBRARY PLASMIDS IN THE CLONES SELECTED.

To identify the proteins which are able to interact with the central region of parkin, the fusion library plasmids contained in the yeasts which were selected during the double-hybrid screening were extracted. To be able to obtain a large amount thereof, this isolation calls for a prior transformation of *E. coli* with an extract of DNA from the positive yeast strains. As the library plasmid which is contained in this extract is a yeast/*E. coli* shuttle plasmid, it can easily replicate in the bacterium. The library plasmid was selected by complementing the auxotrophic HB101 bacterium for leucine, on leucine-lacking medium.

The plasmid DNAs from the bacterial colonies which are obtained after transformation with extracts of DNA from yeasts were analyzed by digestion with restriction enzymes and separation of the DNA fragments on agarose gel. Among the 115 clones analyzed, one clone containing a library plasmid, which showed a different profile from the others, was obtained. This plasmid, termed pGAD-Ly111b, was studied more precisely.

EXAMPLE 4: DETERMINATION OF THE SEQUENCE OF THE INSERT CONTAINED IN THE PLASMID IDENTIFIED.

Sequencing of the insert contained in the plasmid identified was carried out, firstly, using the oligonucleotide SEQ ID NO: 7, which is complementary to the sequence GAL4TA, close to the *EcoRI* site of insertion of the lymphocyte cDNA library; then, secondly, using the oligonucleotides SEQ ID NO: 8 to SEQ ID NO: 11, which correspond to the sequence of the insert which is obtained during the course of the sequencing. The sequence obtained is presented on the sequence SEQ ID NO: 1. The protein thus identified was referred to as PAP1 (Parkin-Associated Protein 1).

Comparison of the sequence of this insert with the sequences which are contained in the GENBank and EMBL (European Molecular Biology Lab)

databases showed a homology of 25% at the protein level with various members of the synaptotagmin family. The synaptotagmins are part of a family of membrane proteins which are encoded by at least eleven different genes, which are expressed in the brain and other tissues. They contain a single transmembrane domain and two calcium-regulated domains which are termed C_2. It is in this domain that the homology between the synaptotagmins and the PAP1 protein is found. No other significant homology was observed.

EXAMPLE 5: ANALYSIS OF THE SPECIFICITY OF INTERACTION BETWEEN THE CENTRAL REGION OF PARKIN AND THE PAP1 PROTEIN.

To determine the specificity of interaction between the fragment corresponding to the PAP1 protein and the central region of parkin, a two-hybrid test for specific interaction with other nonrelevant proteins was carried out. To carry out this test, we transformed strain L40 with the control plasmids plex9-cAPP or pLex9-HaRasVal12, in place of the plasmid pLex9-parkin (135-290), which respectively encode the cytoplasmic domain of the APP or the HaRasVal12 protein, which are fused to the LexA DNA binding domain, and with the plasmid isolated during the screening of the two-hybrid library. A β-Gal activity assay was carried out on the cells which were transformed with the various plasmids, to determine a protein-protein interaction. According to the result of the assay, only the yeasts which were transformed with the plasmid which was isolated during the screening of the two-hybrid library, and with the plasmid pLex9-parkin (135-290), had a β-Gal+ activity, which thus shows an interaction between the central region of parkin and the PAP1 protein. This interaction thus turns out to be specific, since this fragment of PAP1 does not seem to interact with the cAPP or HaRasVal12 proteins.

These results thus show the existence of a novel protein, referred to as PAP1, which is capable of interacting specifically with parkin. This protein, which is related to the synaptotagmins, shows no significant homology with

known proteins, and can be used in therapeutic or diagnostic applications, for producing antibodies, probes or peptides, or for screening active molecules.

EXAMPLE 6: CLONING OF THE PAP1 GENE FROM A HUMAN LUNG DNA LIBRARY

In order to identify the complete sequence of the human PAP1 gene and characterize the existence of variant forms, two elongation approaches were carried out from the sequence SEQ ID NO: 1. Two sequences were thus obtained, of 1644 bp and 1646 bp respectively, comprising an elongation of 330 bp as compared to the sequence SEQ ID NO: 1. Nonetheless, analysis of these sequences showed differences in the consensus region, which were apparent after translation. Thus an ORF of 420aa is obtained in one case and an ORF of 230aa with the other sequence. The protein sequence obtained was compared with the known sequences and revealed a 24% homology over the 293 amino acids that overlap with the human synaptogamin 1 (p65)(p21579). The function of the synaptogamin 1 can be a regulating role in the membrane interactions which occur during the synaptic vesicle traffic in the area of the synapse. The synaptogamin binds the acidic phospholipids with a certain specificity. Moreover, a calcium-dependent interaction between the synaptogamin and the activated kinase C protein receptors was reported. The synaptogamin can also bind three other proteins, which are neurexin, syntaxin and ap2. Given the premature and abrupt disappearance of any homology between the sequences identified and the family of synaptogamins, the sequence identified may contain a deletion as compared to the natural sequence. To verify this hypothesis and validate the sequences, a RT-PCR and sequencing experiment was carried out using the 1644 bp sequence. The sequence obtained comprises an ORF of 420aa with a homology with the synaptogamins on the same order.

In an effort to obtain a larger sequence and verify whether the sequence obtained could correspond to a form of splicing, a 5'-RACE elongation experiment was begun at the 3' region of the validated sequence, using the L1 and L2 oligonucleotides on a human lung cDNA preparation.

The results obtained appear in Figure 2 and show the identification of 8 clones corresponding to 6 different 5' terminal ends. Three of these contain a stop codon which interrupts the ORF (clones A12, F2, F12) and clone A3 contains no ORF. The presence of various transcripts was confirmed by RT-PCR and nested RT-PCR (Table 1).

Table 1

RT-PCR	Primary	Secondary PCR U3-L3	Secondary PCR U1-L4	Secondary PCR C-B
U3-L3	170			
A-L4	153		+	
A-L3	Smear	+		
U1-L4	130			
U1-L3	Smear	+		
U1-B	415			+
U2-B	515			+
Expected size		170	130	120

The U3-L3 and C-B primer pairs are specific to the common fragment of the sequence, the A and U1 oligonucleotides are specific to the initial sequence and to clone C11, the L4 oligonucleotide is specific to the initial sequence and the U2 primer is specific to clone A3. A second 5'-RACE was carried out with oligonucleotides L3 and L7 located in the common region of the different clones (Figure 2). The results obtained appear in Figures 3 and 4. The presence of different transcripts was confirmed by RT-PCR and nested RT-PCR (Table 2).

Table 2

RT-PCR	Result	Secondary PCR C-B	Secondary PCR U3-B	Secondary PCR U5-L7
U4-F	Smear	+	+	+
U5-F	Smear	+	+	+
U3-F	1550 bp	+	+	
Expected Size (bp)		120	385	530

The primer and oligonucleotide sequence is provided in Tables 3 and 4 (SEQ ID NO: 16-37).

Table 3

				SEQ ID
LY111_U4	GCAGTTCTGCCTGTTCATC	23	to 41	16
LY111_U5	TTCAAAACACAGAGGAGGAG	319	to 338	17
LY111_U3	GAATTTGGTCAGTTTAGAGG	759	to 778	18
LY111_L7	TTCTGGGATTTGGAGAGCTTTTTCAC	851	to 825	19
LY111_L6	TCTGTCTGTCCCACACACTGCC	914	to 892	20
LY111_L3	GACTGGCTCCGTCTCTCTG	928	to 910	21
LY111_C	AAGCAACAGAATCTCCCATCC	1029	to 1049	22
LY111_B	GCATTGTCAAAATTGCCCATC	1147	to 1127	23
LY111_E	AGGCGGAGAAATACGAAGAC	1543	to 1562	24
LY111_D	GCAGAGTGAGACAGCCCTTAAC	1767	to 1746	25
Ly111_L2	CTTCCTCAGGACTGGCGACTTCAG	1811	to 1782	26
Ly111_L1	CAAGCGGTCGTTCATTCCAAAGAG	1934	to 1913	27
LY111_F	AAGAGGAGATAACCCACCAGAG	2288	to 2269	28

Table 4

LY111_A	TCGTAGAGCAGCAGGTCCAAG	14	to 34	46
LY111_U1	AGGGCTGCTGGCTATTTTTC	36	to 55	29
LY111_L4	TAAGAAATGGGTTGTGAAC	148	to 166	30
LY111_C	AAGCAACAGAATCTCCCATCC	1029	to 1049	31
LY111_B	GCATTGTCAAAATTGCCCATC	1147	to 1127	32
LY111_E	AGGCGGAGAAATACGAAGAC	1543	to 1562	33
LY111_D	GCAGAGTGAGACAGCCCTTAAC	1767	to 1746	34
Ly111_L2	CTTCCTCAGGACTGGCGACTTCAG	1811	to 1782	35
Ly111_L1	CAAGCGGTCGTTCATTCCAAAGAG	1934	to 1913	36
LY111_F	AAGAGGAGATAACCCACCAGAG	2288	to 2269	37

All of these results make it possible to validate the consensus sequence which corresponds to the long isoform (Figure 9, SEQ ID NO: 12 and 13) and the short isoform (Figure 10, SEQ ID NO: 14 and 15) of the PAP1 protein which was identified from human lung. This protein is also referred to in the following examples as Ly111. The long isoform is encoded by an ORF of 1833 bp, located at residues 237-2069 of SEQ ID NO: 12 and comprises 610 amino acids. The polyadenylation signal is located from nucleotide 2315. The short isoform is encoded by an ORF of 942 bp, located at residues 429-1370 of SEQ ID NO: 14, and comprises 313 amino acids. The polyadenylation signal is located from nucleotide 1616.

Northern blot experiments were then performed on various human tissues with probes (amplimer CD and E-F) and made it possible to reveal a 6 kb transcript in the muscle, a transcript in the heart (3 kb), as well as a 6 kb transcript in the fetal liver. In addition, Example 7 describes the cloning of a transcript in the human fetal brain. Various homology studies were carried out in different protein databases and the results thereof are presented in Table 5, below.

Table 5

Library	Homology
Genpept116	G5926736 (AB025258) granuphilin-a Identity: 31% (215/679), Homology (POS): 46% (322/679)
	G5926738 (AB025259) granuphilin-b Identity: 31% (150/479), Homology (POS): 47% (230/479)
	G1235722 (D70830) Doc2 beta (homo sapiens) Identity: 25% (74/292), Homology (POS): 43% (127/292)
	G289718 (L15302) Synaptogamin-I Identity: 26% (77/293), Homology (POS): 45% (133/293)
Swissprot	SP:SYTI_CAEEL Synaptogamin I Identity: 26% (77/293), Homology (POS): 45% (133/293)
	SP:SYT2_MOUSE Synaptogamin II Identity: 24% (72/293), Homology (POS): 44% (131/293)

EXAMPLE 7: CLONING OF TWO FULL-LENGTH PAP1 (LY111B) TRANSCRIPTS FROM COMPLEMENTARY HUMAN FETAL BRAIN DNA

In order to confirm the presence of a full-length Ly111b transcript in the human brain, a PCR was performed from complementary DNA taken from human fetal brain (Marathon Ready cDNA, Clontech), using the oligonucleotides LyF1 (AAT GGA AGG GCG TGA CGC, Figure 5, SEQ ID NO: 38) and HA71 (CCT CAC GCC TGC TGC AAC CTG, SEQ ID NO: 39) as primers. A DNA fragment with low representation of approximately two kilobases was amplified. The product of this first PCR served as a matrix for a nested PCR, carried out with oligonucleotides LyEcoF (GCACGAATTC ATG GCC CAA GAA ATA GAT CTG, SEQ ID NO: 40) and HA72 (CTG TCT TCG TAT TTC TCC GCC TTG, SEQ ID NO: 41). The amplified products were digested with the restriction enzymes EcoRI (integrated into the oligonucleotide LyEcoF) and BstEII (Figure 5) and inserted into the expression vector pcDNA3, then their sequence was determined. Analysis of the clone sequences obtained revealed the presence of two potential full-length Ly111b transcripts in the human fetal brain (Figure 5). The first of these transcripts ($Ly111b_{fullA}$) corresponds to the mRNA which was identified in the human lung (Example 6) and encodes a protein of 609 amino acids ($pLy111b_{fullA}$; Figures 5,6, SEQ ID NO: 42-43). The second ($Ly111b_{fullB}$) probably represents an alternative splicing product of a common primary mRNA. In this transcript, which is identical to $Ly111b_{fullA}$, the sequence between nucleotides 752 and 956 of the sequence validated in the human lung is absent (SEQ ID NO: 42). $Ly111b_{fullB}$ thus encodes a protein of 541 amino acids ($pLy111b_{fullB}$) which is identical to $pLy111_{fullA}$, in which, however, the domain included between amino acids 172 and 240 (Figures 5,7, SEQ ID NO: 44-45) comes to be missing. The two proteins $pLy111b_{fullA/fullB}$ integrate into the domain of interaction with the fragment of parkin that comprises amino acids 135 to 290, which were identified in the yeast (initial sequence Ly111b, Figure 5), and can therefore theoretically maintain this interaction.

The pLy111b$_{fullA/fullB}$ proteins belong to the RIM/Rabphiline family

pLy111b$_{fullA/fullB}$ shows a homology with the proteins of the RIM/Rabphiline family (Wang Y. Sugita S & Sudhof TG. The RIM/NIM Family of Neuronal C2 Domain Proteins. *J Biol Chem* (2000) 275.20033-20044) and in particular with the granulophilins (Wang Jie, Takeuchi T, Yokota H & Izumi T. Novel Rabphilin-3-like Protein Associates with Insulin-containing Granules in Pancreatic Beta Calls. *J Biol Chem* (1999) 274, 28542-28548). They are characterized by the presence of a zinc finger domain in the N-terminal part of the two C_2 domains, in the C terminal part (Figures 6 and 7). The zinc finger domain of the proteins from the RIM/Rabphiline family was involved in the interaction with the Rab proteins. These Rab proteins, which bind GTP, are compounds which are essential to the machinery of membrane traffic in the eukaryotic cells. Moreover, it has been described that the C_2 domains of the proteins from the RIM/Rabphiline family can bind membranes by interacting with phospholipids.

Expression of the pLy111b$_{fullA/fullB}$ proteins in the cells of the cos-7 line: co-localization with parkin

The coding sequence of the Ly111b$_{fullA,B}$ transcripts was inserted into the eukaryotic expression vector pcDNA3 in phase with the sequence which encodes a myc N-terminal epitope (pcDNA3-mycLy111b$_{fullA/B}$). Cells from the cos-7 line which are transfected using these vectors produce proteins with an apparent molecular weight of approximately 67 kDa (pcDNA3-mycLy111b$_{fullA}$) and 60 kDa (pcDNA3-mycLy111b$_{fullB}$), which corresponds to the expected molecular weight. These proteins, which were detected via immunolabelling, using an antibody directed against the N-terminal myc epitope, are distributed in the cytoplasm, the extensions and at times the nucleus of the cos-7 line of cells in a non-homogenous, punctate manner (Figure 8a, b, column A). When these proteins are overexpressed with parkin and revealed using the

Asp5 anti-parkin antibody in the cells of line cos-7 (Figure 8a, b, column B) a similar distribution pattern and a co-localization of these proteins can be observed (Figure 8a, b, column C).

BIBLIOGRAPHIC REFERENCES

Abbas, N. *et al.* (1999). A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. *Hum Mol Genet.* **8**, 567-574.

Bartel, P.L. *et al.* (1993). D.A. Hartley Ed, Oxford University Press, 153

Chung, C.T. *et al.* (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution, *Proc. Natl. Acad. Sci. USA.* **86**, 2172-2175.

Fields, S. and **Song,** O. (1989). A novel genetic system to detect protein-protein interactions. *Nature.* **340,** 245-246.

Gietz, R.D. *et al.* (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.,* **20**, 1425.

Hattori, N. *et al.* (1998). Molecular genetic analysis of a novel Parkin gene in Japanese families with autosomal recessive juvenile parkinsonism: evidence for variable homozygous deletions in the Parkin gene in affected individuals. *Ann Neurol.* **6**, 935-941.

Kahn, A. *et al.*, (1991) Thérapie génique: espoirs et limites, *Médecine et Sciences.* **7**, 705-714.

Kitada, T. *et al.* (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature.* **392,** 605-608.

Leroy, E. *et al.* (1998). The ubiquitin pathway in Parkinson's disease. *Nature.* **395**, 451-452.

Lewy, F.H. (1912). in *Handbuch der Neurologie* (Lewandowski, M., ed) pp 920-933, Springer, Berlin.

Lucking, C. *et al.* (1998). Homozygous deletions in parkin gene in European and North African families with autosomal recessive juvenile parkinsonism. *Lancet.* **352,** 1355-1356.

Maniatis, T. *et al.* (1989). Molecular cloning, second edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Morett, E. (1999). A novel transactivation domain in parkin. *Trends Biochem Sci.* **24**, 229-231.

Peytavi, R. *et al.* (1999). HEED, the product of the human homolog of the murine eed gene, binds to the matrix protein of HIV-1. *J. Biol. Chem.* **274,** 1635-1645.

Polymeropoulos, M.H. *et al.* (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*. **276**, 2045-2047.

Sanger, F. *et al.* (1997). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463-5467.

Shimura, H. *et al.* (1999). Immunohistochemical and subcellular localization of Parkin protein: absence of protein in autosomal recessive juvenile parkinsonism patients. *Ann Neurol.* **45**, 668-672.

Sunada, Y. *et al.* (1998), Differential expression of the parkin gene in the human brain and peripheral leukocytes. *Neurosci Lett.* **254**, 180-182.

Vojtek, AB. *et al.,* (1993). Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell.* **74**, 205-214.